A framework for the discovery of retinal biomarkers in Optical Coherence Tomography Angiography (OCTA)

A framework for the discovery of retinal biomarkers in Optical Coherence Tomography Angiography (OCTA)

Ylenia Giarratano[1], Alisa Pavel[2,3,5], Jie Lian[4,5], Rayna Andreeva[5], Alessandro Fontanella[5], Rik Sarkar[5], Laura J. Reid[6], Shareen Forbes[6], Dan Pugh[6], Tariq E. Farrah[6], Neeraj Dhaun[6], Baljean Dhillon[7,8], Tom MacGillivray[8], and Miguel O. Bernabeu[1]

[1] Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK ylenia.giarratano@ed.ac.uk
[2] Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
[3] BioMediTech Institute, Tampere University, Tampere, Finland
[4] Institute of Computer Science, University of St. Gallen, St. Gallen, Switzerland
[5] School of Informatics, University of Edinburgh, Edinburgh, UK
[6] Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK
[7] School of Clinical Sciences, University of Edinburgh, Edinburgh, UK
[8] Centre for Clinical Brain Sciences, University of Edinburgh, UK

Abstract. Recent studies have demonstrated the potential of OCTA retinal imaging for the discovery of biomarkers of vascular disease of the eye and other organs. Furthermore, advances in deep learning have made it possible to train algorithms for the automated detection of such biomarkers. However, two key limitations of this approach are the need for large numbers of labeled images to train the algorithms, which are often not met by the typical single-centre prospective studies in the literature, and the lack of interpretability of the features learned during training. In the current study, we developed a network analysis framework to characterise retinal vasculature where geometric and topological information are exploited to increase the performance of classifiers trained on tens of OCTA images. We demonstrate our approach in two different diseases with a retinal vascular footprint: diabetic retinopathy (DR) and chronic kidney disease (CKD). Our approach enables the discovery of previously unreported retinal vascular morphological differences in DR and CKD, and demonstrate the potential of OCTA for automated disease assessment.

Keywords: Optical coherence tomography angiography · Vascular network · Graph analysis · Retinal biomarkers

1 Introduction

Optical coherence tomography angiography (OCTA) is a fast and efficient imaging modality that allows the visualisation of retinal vasculature at the capillary

level. This technology offers the advantage of visualising *in vivo* the microvasculature without any invasive procedure, emerging as a promising modality to investigate microvascular disease. However, the usefulness of OCTA as a diagnostic tool depends on the availability of accurate and reproducible image quantification metrics to identify retinal microvascular changes. Machine learning (ML), and in particular deep learning (DL) in recent years, has emerged as a promising approach for automated image analysis. Despite the popularity of DL in medical applications, these techniques typically require large amounts of labeled data for training, and a straight forward clinical interpretation of the features learned during training might not be possible. Hence, several existing approaches in the literature for the analysis of OCTA scans have focused on a small set of interpretable candidate biomarkers based on accumulated clinical knowledge. In diabetic retinopathy, metrics related to the morphology of the foveal avascular zone (FAZ), the central part of the retina responsible for the sharpest vision, and vascular-based metrics such as vessel density and capillary nonperfusion have been used as biomarkers to investigate disease progression [6, 8, 10, 15]. In addition, vessel density has been considered in Alzheimer's disease [18, 21] and chronic kidney disease [17], highlighting the potential of retinal OCTA imaging to investigate disease of other organs, an area of increasing interest [19]. Albeit promising, these phenotypes are not representative of the full spectrum of retinal vascular morphometric characteristics that could be exploited for diagnosis. Hence, novel approaches to biomarker discovery are urgently needed. Furthermore, automated classification of diabetic retinopathy severity has been recently explored and support vector machine classification models, based on vessel density, vessel caliber, and FAZ measurements, have been proposed in [1, 11]. Transfer learning was recently used to reduce the volume of data required to train DL models for the prediction of diabetic retinopathy patient status from OCTA images [7] without the need of feature engineering. However, such approach has not been able to provide to date interpretability of results.

In the current study, we propose a fully automated approach to the classification of OCTA images according to disease status based on interpretable retinal vascular features (i.e., quantifications of vascular characteristics). Our framework enables: hypothesis-free discovery of new retinal biomarkers of disease; development of ML classifiers based on modestly sized image datasets without compromising features interpretability. We propose new microvascular metrics based on geometrical and topological properties of the graph representation of the vasculature and we show applications in two case studies: diabetic retinopathy (DR) and chronic kidney disease (CKD). In the DR substudy, we replicate previous findings of changes in vessel morphology and vessel density [4, 6] and discover new discriminative topological features. In the CKD substudy, we discover previously unreported structural and functional changes, and we are the first to report automated classification of CKD patient status based on OCTA retinal imaging. Finally, we show our DR and CKD classifiers outperform or achieve comparable performances of state-of-the-art DL approaches even when considering transfer learning, with the added advantage of features interpretability.

2 Methods

2.1 Vascular Graph construction

Images of the left and right eye of the study participants are obtained using a commercial OCTA device (RTVue XR Avanti; Optovue). Only the superficial layer (from the internal limiting membrane layer (ILM) to the inner plexiform layer (IPL)), are considered for all metrics. Superficial and deep layers (from the IPL to the bottom of the outer plexiform layer) with 3×3 mm field of view are used to compute FAZ metrics (Figure 1A). Input of our approach is the binary mask of the retinal plexus of interest (Figure 1B). Firstly, we apply automated segmentation by using a U-Net architecture as described in [2]. Morphological thinning is then performed on the binary image to obtain a one pixelwise wide skeleton of the vasculature. In order to construct the graph and preserve the morphological structure of the network, we use a map that, considering the skeletonized image, S, of $n \times n$ pixels, for each row i and column j, the element $s_{ij} \in \{0, 1\}$ is associated to the coordinates (i, j). Vertices of our graph are then found as coordinates of the white pixel of the binary mask. To construct edges, we analyse the neighborhood of size 2×2, drawing firstly vertical and horizontal connections, and in absence of links, we explore connections along the diagonals. This procedure allows us to find edges that preserve the morphological structure of the network avoiding over connectivity. Finally, to each node, we assign two attributes: coordinates and radius. The latter is used to compute edge attribute thickness as the average radius of its end points, and keep track of vessel width. Radii are calculated starting from the original image, by computing vessel boundaries and their Euclidean distance from the pixel centre-line representing our vertex. The final network is rescaled according to the original pixel size (Figure 1C). Considering the anatomy of the vasculature, we assume small disconnected components as segmentation artifacts, and only the largest connected component is used in our analysis.

2.2 Graph simplification

Due to the size of the images, the final vascular network is highly dense in the number of nodes and edges. To compute morphological metrics that require to explore all the nodes in the network, which can be a very time-consuming procedure, we adopted a simplified version of the graph that preserves the vasculature structure. Briefly, let G = (V, E) be our initial graph and $e_A, e_B \in E$ two adjacent edges of starting and end points (u, v) and (v, w), respectively. If e_A and e_B lay on the same line then remove e_A and e_B, and create a new edge, $e_{AB} = (u, w)$. This simplification is used, for example, to speed up the process of finding the intercapillary space, since it reduces computational time, without affecting the morphology of the network (see Figure 1C,E).

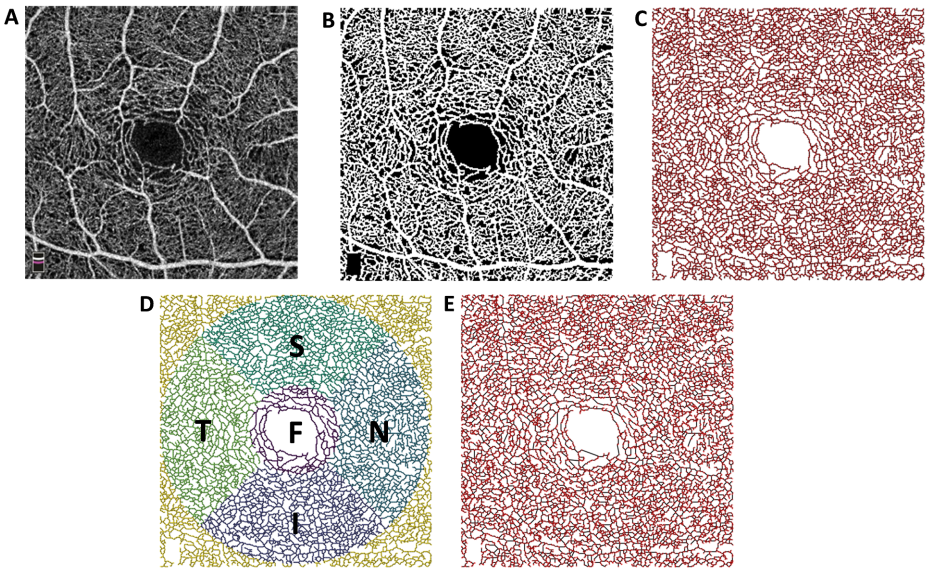

Fig. 1. (A) Original image (Superficial+Deep layers). (B) Binary image obtained using U-Net. (C) Graph representation of the vascular network. (D) Regions of interest in the network: foveal (**F**), superior (**S**), nasal (**N**), inferior (**I**), temporal (**T**). (E) Simplified graph (nodes in red, edges in black).

2.3 Feature extraction

Our graph representation of the vascular network allows us to exploit many vascular interpretable features that have never been used to investigate OCTA images before. We compute the following features for each of the regions of interest in the retinal plexus (foveal, superior, nasal, inferior, and temporal, Figure 1D). In the case of distributions, mean, median, variance, skewness, and kurtosis are reported.

Graph-based features An overview of the vascular network is obtained by computing basic graph metrics. Graph density measures the sparseness of the network, average clustering coefficient is used to measure how closely connected the nodes in the network are. Graph diameter describing the maximum eccentricity, i.e. the maximum distance between any two nodes in the network and the graph radius, describing its minimum, are calculated to understand the span of the network. Finally, edges thickness is used to investigate vessels widening and thinning possibly associated with diseases.

Coordinate-based features These metrics are related to the position of each node in the network. They mainly capture the morphology and size of portions of the graph. Among these metrics, we find common retinal measures related to the

foveal avascular zone (FAZ), a crucial part in the retina that has been observed widening in people with diabetic conditions [15]. We are interested in FAZ area, perimeter, and shape. FAZ area (A) and perimeter (P) are computed by considering all the points making up its boundary as vertices of a polygon. Based on this measurement FAZ circularity is defined as $C = 4\pi A/P^2$. To characterise FAZ shape we use the acircularity index, computed as $P/(2\pi R)$, where R is the radius of a circle of size equal to the FAZ area, and the axis ratio, calculated by fitting an ellipse in the FAZ and then computing major and minor axis. Supposing FAZ boundary, B, describing a path on which an object is moving, and each node of coordinate (x_i, y_i) in B as an observation of the object taken in constant time intervals, we can then compute the speed of the object, its curvature (the deviation of a curve from a straight line as $k = |x'y'' - y'x''|/(x'^2 + y'^2)^{3/2})$, and the number of turning points to estimate boundary smoothness.
Finally, vessel tortuosity has been previously associated with diabetes. We define the tortuosity as the ratio of the length of the path between two branching points against the Euclidean distance between its end points.

Flow-based features Characterisation of the resistance to blood flow of a given vessel can be done by using Poiseuille's law defined as $R(l, r) = \Delta P/Q = 8\eta l/\pi r^4$, where ΔP is the pressure difference between the two ends of the vascular segment and Q is the flow rate. Given the length (l) and radius (r) of the vessel and assuming constant viscosity of blood ($\eta = 2.084 \times 10^{-3} Pa \cdot s$), we can compute the resistance at each vessel segment.
Another measurement that we extract is the area to flow capacity ratio. Tissue is supplied by oxygen and nutrients carried in the vessels, considering the size of the intercapillary space (a face in the graph), the radius, and length of the vessels that enclose it, we can estimate the ratio between tissue area and flow capacity (under the assumption of vessels cylindrical in shape).

Topology-based features The retinal microvasculature is a complex and intertwined network, the investigation of changes underlying its loopy structure can reveal valuable insight into disease detection and progression. As suggested in [3] and [9] biological networks can be mapped into binary tree structures. The mapping implementation is based on the algorithm described in [9]: starting from the retinal loopy graph (containing only loops and therefore free of nodes of degree one), we find the area enclosed by edges, called *faces*, representing the leaves of the tree. At each step, we remove the shortest edge in the face boundary so that small loops are merged into bigger loops. To find the initial faces, we use the doubly connected edge list (DCEL) and the algorithm described in [12]. To ensure a binary tree root a pseudo loop surrounding the initial network is created. The binary tree mapping allows us to exploit new metrics, such as tree depth, indicating how homogeneous the loops are distributed over the network structure, tree leaves, showing the number of loops, tree asymmetry, how much the structure deviates from a perfect binary tree. To characterise the shape, we can use the number of exterior and interior edges, where an external edge has a

bifurcation at its upstream end, tree altitude and total exterior path length [20]. Strahler-branching ratio is calculated by assigning a hierarchical ordering to the edges of the binary tree. We can then, classify each external edge as external-external or external-internal based on the Strahler order of the sibling edge [14]. Topological pattern comparison can be done by introducing graphlets up to size 4: small subgraphs whose distribution can elucidate repeated motives in the network. To reduce computational time, graphlets distribution is estimated based on a sample distribution. Finally we use random walks to characterise the average path length in the network. We select 1000 walks of length 300, where all the walks start from a randomly selected node, and compare it to the euclidean distance between start and end node.

2.4 Demographics and statistical analysis

We considered three groups of participants: 26 diabetic subjects with and without diabetic retinopathy (12 DR and 14 NoDR, respectively), 25 subjects that suffer from chronic kidney disease (CKD), and 25 age- and gender-matched healthy subjects (Controls). For each participant, only one eye was considered for our analysis. All the included images were free from major observable artifacts such as vertical and horizontal line distortions. Each patient scan was used by our framework to extract meaningful morphological and topological metrics. Statistical analysis was then performed to investigate features that are significantly different across groups and possibly associated with disease status. Shapiro-Wilk test was used to assess features normally distributed. In three groups comparison (DR, NoDR, and Controls) we performed one-way analysis of variance (ANOVA), in the case of variables normally distributed, and non-parametric Kruskal-Wallis test otherwise. For one versus one comparisons, we used t-tests for the normally distributed features, and Mann–Whitney test in the failure of Shapiro-Wilk test. In the case of multiple comparison, Bonferroni correction was applied.
After removing highly correlated metrics, we performed feature selection within a ten-fold cross-validation. In DR study, 30 features with the highest mutual information were selected before using Random Forest as classification model. In the CKD study, only the features that are statistically significant in each fold were used before applying support vector machines (SVM). Finally, we provided comparisons of our models with state-of-the-art DL approaches to patient classification. A VGG16 architecture with transfer learning was used as described in [7] to classify the same OCTA images. Ten-fold cross-validation and data augmentation was used in both sub-studies. Sensitivity, specificity, accuracy, and area under the curve (AUC) were reported for performance evaluation.

3 Results

Diabetic cohort Among the 189 statistically significant features, 79 were graph-based, 28 coordinate-based, 67 topology-based, and 15 flow-based. From

these features, vessel skeleton density and the enlargement of the intercapillary spaces have been already reported [5, 16]. Figure 2A-C displays three features of interest: a) number of nodes, which is equivalent to the previously reported vessel skeleton density metric (since nodes represent pixels in the skeleton); b) mean size of the intercapillary spaces, a candidate for simple clinical inspection; and c) the median circularity of the intercapillary space in the nasal segment, which could distinguish between Controls and DR, and between Controls and NoDR, highlighting the importance of shape analysis. Interestingly, these three features showed a monotonic decrease/increase going from Controls to NoDR and finally DR, highlighting that the network properties of NoDR appear as an intermediate stage between Controls and DR. Random Forest provided the highest AUC (0.84) in the classification of the three groups, outperforming VGG16 with transfer learning and data augmentation (AUC= 0.79) (Table 1).

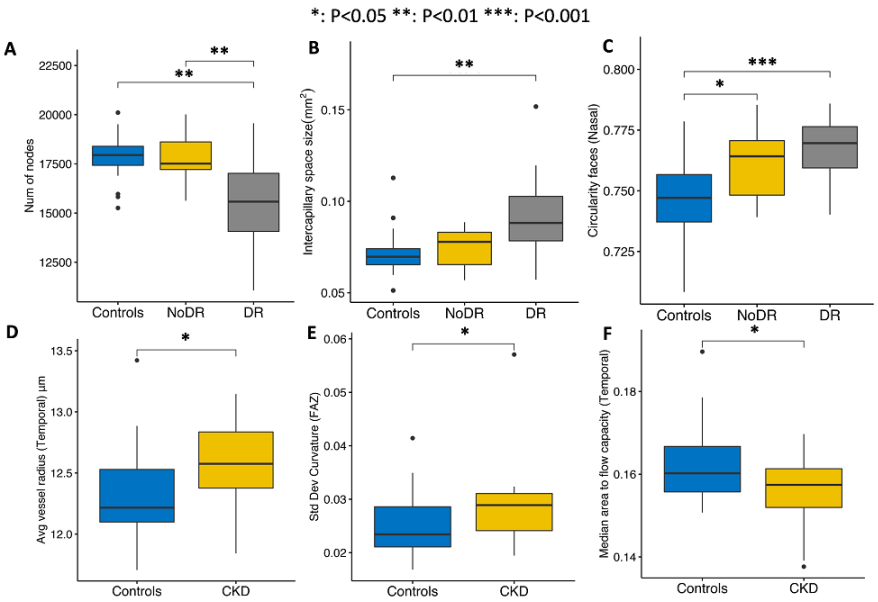

Fig. 2. Statistical significant features in diabetic (DR, NoDR) and CKD groups compared with Controls.

CKD cohort We found 43 statistically significant features, including 10 graph-based metrics, 20 coordinate-based metrics, 10 topology-based metrics, and 3 flow-based metrics. Our results showed differences in previously unreported morphological and flow metrics. The CKD group shows larger values of vessel radius in the temporal region of the retina, as well as a high imbalance in the curvature measurements of the FAZ, indicating high variability of this parameter in the

Table 1. Table of classification performances in DR study

	Controls		NoDR		DR	
	Our approach	**VGG16**	**Our approach**	**VGG16**	**Our approach**	**VGG16**
ACC (mean ± SE)	**0.80±0.07**	0.78 ± 0.05	**0.75 ± 0.08**	0.72 ± 0.04	**0.82 ± 0.06**	0.77 ± 0.04
SEN (mean ± SE)	0.79 ± 0.08	**0.90 ± 0.05**	**0.30 ± 0.10**	0.20 ± 0.13	**0.55 ± 0.19**	**0.55 ± 0.11**
SPE (mean ± SE)	**0.77 ± 0.08**	0.67 ± 0.11	**0.88 ± 0.05**	**0.88± 0.05**	**0.88 ± 0.05**	0.86 ± 0.05
AUC (mean ± SD)	**0.90 ± 0.11**	**0.90±0.15**	**0.67 ± 0.37**	**0.67± 0.28**	**0.81 ± 0.12**	0.75 ± 0.22

disease group. Moreover, the newly introduced area to flow capacity has been found lower in the temporal area of the CKD group underlining the strong connection between structural and functional changes (2D-F). In contrast with our diabetic cohort (and previous DR studies), common measurements such as vessel density and size of intercapillary spaces did not differ between CKD and controls, highlighting that the retinal vascular changes in CKD could be perceived clinically as much more subtle. This is also confirmed by classification performances. Both SVM and VGG16 performed poorly in the classification task (Table 2), achieving the same AUC equals to 0.62.

Table 2. Table of classification performances in CKD study

	Controls		CKD	
	Our approach	**VGG16**	**Our approach**	**VGG16**
ACC	**0.52± 0.09**	0.48 ± 0.08	**0.52± 0.09**	0.48 ± 0.08
SEN	0.58 ± 0.14	**0.61 ± 0.09**	**0.47± 0.15**	0.38 ± 0.12
SPE	**0.47± 0.15**	0.38 ± 0.12	0.58 ± 0.14	**0.61 ± 0.09**
AUC	0.62 ± 0.29	0.62 ± 0.37	0.62 ± 0.29	0.62 ± 0.37

4 Discussion and conclusions

With the increasing interest in mining the retinal landscape for biomarkers of both eye and systemic disease, which some authors have termed oculomics [19], OCTA has become a key modality for the study of the retinal microvascular system *in vivo*. Previous studies have shown associations between changes in the retinal microvascular structure and conditions as diverse as diabetic retinopathy, Alzheimer's disease, and chronic kidney disease. Early identification of these alterations can contribute to timely target patients at risk and effectively monitor disease progression. Quantifying these changes by matching selected vascular features may provide a non-invasive alternative to characterise vascular dysfunctions and gain valuable insights for clinical practice. In this study, we propose a novel network analysis framework for hypothesis-free discovery of new retinal biomarkers of disease and the development of machine learning classifiers based

on modestly sized image datasets without compromising on the interpretability of the classification features. These geometric and topological features may elucidate vascular alterations and unravel clinically relevant measurements. We demonstrate our approach in two different diseases with a retinal vascular footprint: diabetic retinopathy and chronic kidney disease. In the diabetic cohort, we are capable of replicating previous findings of structural alterations associated with disease. In both sub-studies, we uncover a large number of previously unreported biomarkers. Furthermore, our automated image classification approach outperforms or achieves comparable results of state-of-the-art deep learning approaches with the added advantage of feature interpretability. Our results confirm that automated classification of chronic kidney disease status based on OCTA imaging is a more challenging task. Future work should investigate the development of more advanced feature selection approaches for model construction that can leverage the large number of biomarkers discovered without suffering from what is known as peaking phenomenon (too large set of features leads to poorer performances than a small set of selected features [13]), and validate our results on larger cohorts to support statistical analysis. Our methodology lays the foundation of a novel prospective approach for retinal microvascular analysis that can realise the full diagnostic potential of OCTA imaging.

References

1. Alam, M., Zhang, Y., Lim, J.I., Chan, R.V., Yang, M., Yao, X.: Quantitative Optical Coherence Tomography Angiography Features for Objective Classification and Staging of Diabetic Retinopathy. Retina (Philadelphia, Pa.) **40**(2), 322–332 (2020)
2. Giarratano, Y., Bianchi, E., Gray, C., Morris, A., MacGillivray, T., Dhillon, B., Bernabeu, M.O.: Automated Segmentation of Optical Coherence Tomography Angiography Images: Benchmark Data and Clinically Relevant Metrics. arXiv:1912.09978v2 (2020)
3. Katifori, E., Magnasco, M.O.: Quantifying loopy network architectures. PLoS ONE **7**(6) (2012)
4. Khadamy, J., Aghdam, K., Falavarjani, K.: An update on optical coherence tomography angiography in diabetic retinopathy. Journal of Ophthalmic & Vision Research **13**, 487 (2018)
5. Kim, A.Y., Chu, Z., Shahidzadeh, A., Wang, R.K., Puliafito, C.A., Kashani, A.H.: Quantifying microvascular density and morphology in diabetic retinopathy using spectral-domain optical coherence tomography angiography. Investigative Ophthalmology and Visual Science **57**(9), OCT362–OCT370 (2016)
6. Krawitz, B.D., Mo, S., Geyman, L.S., Agemy, S.A., Scripsema, N.K., Garcia, P.M., Chui, T.Y., Rosen, R.B.: Acircularity index and axis ratio of the foveal avascular zone in diabetic eyes and healthy controls measured by optical coherence tomography angiography. Vision Research **139**, 177–186 (2017)
7. Le, D., Alam, M.N., Lim, J.I., Chan, R.V.P., Yao, X.: Deep learning for objective OCTA detection of diabetic retinopathy. In: Manns, F., Ho, A., Söderberg, P.G. (eds.) Ophthalmic Technologies XXX. vol. 11218, pp. 98 – 103. International Society for Optics and Photonics, SPIE (2020)

8. Mastropasqua, R., Toto, L., Mastropasqua, A., Aloia, R., De Nicola, C., Mattei, P.A., Di Marzio, G., Di Nicola, M., Di Antonio, L.: Foveal avascular zone area and parafoveal vessel density measurements in different stages of diabetic retinopathy by optical coherence tomography angiography. International Journal of Ophthalmology **10**(10), 1545–1551 (2017)
9. Modes, C.D., Magnasco, M.O., Katifori, E.: Extracting hidden hierarchies in 3D distribution networks. Physical Review X **6**(3), 1–12 (2016)
10. Nesper, P.L., Roberts, P.K., Onishi, A.C., Chai, H., Liu, L., Jampol, L.M., Fawzi, A.A.: Quantifying Microvascular Abnormalities With Increasing Severity of Diabetic Retinopathy Using Optical Coherence Tomography Angiography. Investigative ophthalmology & visual science **58**(6), BIO307–BIO315 (2017)
11. Sandhu, H.S., Eladawi, N., Elmogy, M., Keynton, R., Helmy, O., Schaal, S., El-Baz, A.: Automated diabetic retinopathy detection using optical coherence tomography angiography: A pilot study. British Journal of Ophthalmology **102**(11), 1564–1569 (2018)
12. Schneider, S., Sbalzarini, I.F.: Finding faces in a planar embedding of a graph. Tech. rep., MOSAIC Group, MPI-CBG (2015)
13. Sima, C., Dougherty, E.R.: The peaking phenomenon in the presence of feature-selection. Pattern Recognition Letters **29**(11), 1667–1674 (2008)
14. Smart, J.S.: The analysis of drainage network composition. Earth Surface Processes **3**(2), 129–170 (1978)
15. Takase, N., Nozaki, M., Kato, A., Ozeki, H., Yoshida, M., Ogura, Y.: Enlargement of foveal avascular zone in diabetic eyes evaluated by en face optical coherence tomography angiography. Retina **35**(11), 2377–2383 (2015)
16. Tang, F.Y., Ng, D.S., Lam, A., Luk, F., Wong, R., Chan, C., Mohamed, S., Fong, A., Lok, J., Tso, T., Lai, F., Brelen, M., Wong, T.Y., Tham, C.C., Cheung, C.Y.: Determinants of Quantitative Optical Coherence Tomography Angiography Metrics in Patients with Diabetes. Scientific Reports **7**(1), 1–10 (2017)
17. Vadalà, M., Castellucci, M., Guarrasi, G., Terrasi, M., La Blasca, T., Mulè, G.: Retinal and choroidal vasculature changes associated with chronic kidney disease. Graefe's Archive for Clinical and Experimental Ophthalmology **257**(8), 1687–1698 (2019)
18. Van De Kreeke, J.A., Nguyen, H.T., Konijnenberg, E., Tomassen, J., Den Braber, A., Ten Kate, M., Yaqub, M., Van Berckel, B., Lammertsma, A.A., Boomsma, D.I., Tan, S.H., Verbraak, F., Visser, P.J.: Optical coherence tomography angiography in preclinical Alzheimer's disease. British Journal of Ophthalmology pp. 157–161 (2019)
19. Wagner, S.K., Fu, D.J., Faes, L., Liu, X., Huemer, J., Khalid, H., Ferraz, D., Korot, E., Kelly, C., Balaskas, K., Denniston, A.K., Keane, P.A.: Insights into Systemic Disease through Retinal Imaging-Based Oculomics. Translational Vision Science & Technology **9**(2), 6 (2020)
20. Werner, C., Smart, J.S.: Some New Methods of Topologic Classification of Channel Networks. Geographical Analysis **5**(4), 271–295 (1973)
21. Yoon, S.P., Grewal, D.S., Thompson, A.C., Polascik, B.W., Dunn, C., Burke, J.R., Fekrat, S.: Retinal Microvascular and Neurodegenerative Changes in Alzheimer's Disease and Mild Cognitive Impairment Compared with Control Participants. Ophthalmology Retina **3**(6), 489–499 (2019)